

Aperture shape optimization in intensity-modulated radiation therapy planning*

Liyuan Zhang,^{1,†} Zhiguo Gui,¹ Pengcheng Zhang,¹ and Jie Yang²

¹Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan, 030051, China

²School of Medicine Management, Shanxi University of Chinese Medicine, Taiyuan, 030619, China

The gradient element of the aperture gradient map is utilized directly to generate the aperture shape without modulation. This process can be likened to choosing the direction of negative gradient descent for the generic aperture shape optimization. The negative-gradient descent direction is more suitable under local conditions and has a slow convergence rate. To overcome these limitations, this study introduced conjugate gradients into aperture shape optimization based on gradient modulation. First, the aperture gradient map of the current beam was obtained for the proposed aperture shape optimization method, and the gradients of the aperture gradient map were modulated using conjugate gradients to form a modulated gradient map. The aperture shape was generated based on the modulated gradient map. The proposed optimization method does not change the optimal solution of the original optimization problem but changes the iterative search direction when generating the aperture shape. The performance of the proposed method was verified using cases of head and neck cancer, and prostate cancer. The optimization results indicate that the proposed optimization method better protects the organs at risk and rapidly reduces the objective function value by ensuring a similar dose distribution to the planning target volume. Compared to the contrasting methods, the normal tissue complication probability obtained by the proposed optimization method decreased by up to 4.61%, and the optimization time of the proposed method decreased by 5.26% on average for ten cancer cases. The effectiveness and acceleration of the proposed method were verified through comparative experiments. According to the comparative experiments, the results indicate that the proposed optimization method is more suitable for clinical applications. It is feasible for the aperture shape optimization involving the proposed method.

Keywords: Aperture shape, Column generation, Conjugate gradient, Gradient modulation, Direct aperture optimization

I. INTRODUCTION

Direct aperture optimization (DAO) [1] for intensity-modulated radiation therapy (IMRT) can be achieved by using approaches such as stochastic search, local gradient-based methods, and column generation. A stochastic search randomly moves the multileaf collimator (MLC) leaves to either side in small increments from their current position [2, 3]. If the motion of the leaves improves the objective function, the current position is updated. Otherwise, the position change is accepted with a certain probability to avoid the local optima. The random nature of stochastic search makes it inefficient for aperture shape optimization. In the local-gradient-based method, the positions of the MLC leaves are the optimization variables in the objective function of the optimization problem [4]. Because this method uses the local gradient of positions to generate an aperture shape, it easily reaches the local optima. In practice, this method largely depends on an appropriate initial solution. An alternative is column generation, which notably differs from the other two methods [5, 6] that no initial aperture shape is required, and the gradient information is not local, as the network flow is constructed using the gradient map of the entire aperture to obtain a deliverable aperture shape. Unlike the two methods mentioned above, the

optimization of column generation does not depend on the suitability of the initial solution. In this study, an improved method based on generic column generation developed.

In one iteration of the generic method, the pricing problem is solved first, followed by the master problem. To solve the pricing problem, a network flow is constructed using the aperture gradient map. It is used to solve the shortest path problem and obtain a deliverable aperture shape. This process is equivalent to choosing a suitable descent direction for the objective function of the optimization problem. If the gradient element of the gradient map is not modulated, the process is equivalent to selecting the steepest descent direction for optimization.

The steepest descent method has the advantage of low computational cost and can converge from any initial point to a local minimum. However, this method typically exhibits a sawtooth effect in the region near the minimum value. Newton's method achieves extremely fast convergence near the optimum, but is computationally expensive. Quasi-Newton methods avoid the explicit matrix required in conventional methods; however, the computation remains highly complex. The conjugate gradient method is an effective substitute because it has a comparable computational cost and converges faster than the steepest descent method. The computational complexity of the conjugate gradient method is less than that of the Newton's and quasi-Newton methods. The conjugate gradient method is particularly suitable for solving large-scale optimization problems and is widely used in economics, engineering, physics, and other fields [7–9]. In IMRT, the conjugate gradient has been used to optimize the weights of beams (apertures) [2, 10] and study the performance of column generation [11].

* Supported by the Natural Science Foundation of Shanxi Province under Grant (No. 20210302124403), the Research Project Supported by Shanxi Scholarship Council of China under Grant (No. 2021-111), and the Science and Technology Innovation Project of Colleges and Universities in Shanxi Province under Grant (No. 2022L353)

† Corresponding author, E-mail: good0806@sina.cn, Tel.: +86-18734914306.

To overcome the drawbacks of aperture shape optimization based on the negative-gradient descent direction, this study introduced conjugate gradients into aperture shape optimization. The search direction containing the conjugate gradient information was constructed using the original gradient of the gradient map to efficiently generate the aperture shape.

II. METHODS

In this study, based on column generation, the aperture shape search direction containing the conjugate gradient components was constructed to overcome the problem of slow convergence in generating an aperture shape directly by using the gradient map without gradient modulation.

A. Dose calculations

During radiotherapy, the patient is irradiated with a pre-defined beam set denoted by B . Each beam in this study consists of m rows and n columns of beamlets, with each beamlet size being $1 \times 1 \text{ cm}^2$. The set of all generated deliverable apertures is denoted by K , and the weight of the aperture κ is y_κ . The beamlets in set A_κ are delivered to the patient via aperture κ , and the treatment involves S structures, where each structure s ($s = 1, \dots, S$) comprises v_s voxels. The element in the deposition matrix is the deposition coefficient W_{ijs} , which represents the dose received by j ($j = 1, \dots, v_s$) in structure s from the beamlet i ($i \in A_\kappa$) allowed to pass through the aperture κ at unit intensity. The dose D_{js} can be expressed as follows:

$$D_{js} = \sum_{\kappa \in K} \left(\sum_{i \in A_\kappa} W_{ijs} \right) y_\kappa. \quad (1)$$

B. Column generation

In this study, the optimization problem is constructed as

$$\min F(D) = \min \sum_{s=1}^S \sum_{\xi=1}^{N_s} F_{\xi s}(D_s), \quad (2)$$

$$\text{s.t.} \sum_{\kappa \in K} \left(\sum_{i \in A_\kappa} W_{ijs} \right) y_\kappa = D_{js}, \quad (3)$$

$$y_\kappa \geq 0, \kappa \in K. \quad (4)$$

$F(D)$ is the objective function of the optimization problem. For s , N_s sub-objective functions $F_{\xi s}(D_s)$ are used to constrain the received dose D_s . An excellent description of column generation was provided by Romeijn et al. [5]. In one iteration, a new aperture shape is generated by solving the pricing problem and is accepted into the apertures set.

The weights of the apertures set are optimized as the master problem. The limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm for bound-constrained optimization [12–14] was employed to optimize the weights of the apertures generated in this study. The optimization was terminated when either the treatment plan met the requirements of the planner or the iteration number reached its limit.

When an aperture is generated, its cost should be calculated [5]. The pricing problem is

$$\min_{\kappa \in K} \sum_{i \in A_\kappa} \left(\sum_{s=1}^S \sum_{j=1}^{v_s} W_{ijs} \pi_{js} \right), \quad (5)$$

where $\pi_{js} = \frac{\partial F_{js}(D_{js})}{\partial D_{js}}$ is the Lagrange multiplier under the Karush–Kuhn–Tucker (KKT) condition. The interdigitation of the MLC was not allowed for the proposed optimization method, and the network flow could be employed to solve the pricing problem with this constraint. For beam $l \in B$, in the r ($r = 1, \dots, m$) row of the beamlets, c_1 ($c_1 = 0, \dots, n$) is used to mark the last beamlet of row r covered by the left leaf, and c_2 ($c_2 = 1, \dots, n+1$) is used to mark the first beamlet of row r covered by the right leaf. According to Eq. (5), the cost of (r, c_1, c_2) is the sum of the gradient elements not covered by the leaves in the gradient map.

Therefore, the radiation beam was decomposed into rectangular beamlets in this study. Each beamlet gradient was calculated and the gradient of the corresponding beamlet was arranged according to the position of the beamlet in the beam, forming the aperture gradient map. An aperture shape that can improve the objective function was generated according to the gradient map. This generated aperture shape was equivalent to the negative-gradient descent direction for the optimization process. The search was performed along the negative-gradient descent direction, which rapidly reduced the function value. However, this did not indicate that the convergence speed of the steepest descent method was high. The sawtooth effect implies that the search direction of a negative gradient is not necessarily the fastest descent direction in the global range.

C. Conjugate gradient modulation

Compared with Newton's and quasi-Newton methods, the conjugate gradient method uses simpler calculations, and its convergence speed is higher than that of the steepest descent method. To speed up the optimization process and improve the optimization quality of column generation, the conjugate gradient method was used to modulate the gradient to generate the aperture shape.

1. Nonlinear conjugate gradient method

Hestenes et al. [15] proposed a linear conjugate gradient method for solving linear equations, whereas Fletcher and Reeves [16] proposed a nonlinear conjugate gradient method

for minimizing general functions. These two methods have been subsequently improved. In the $(k+1)$ th iteration of the generic process of the nonlinear conjugate gradient method, the iteration format is $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$. The conjugate gradient direction \mathbf{d}_k is updated as follows:

$$\mathbf{d}_k = \begin{cases} -\mathbf{g}_k, & k = 1 \\ -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}, & k > 1 \end{cases} \quad (6)$$

where \mathbf{x} denotes the independent variable of the objective function, \mathbf{d} denotes the search direction (i.e., conjugate gradient direction), and \mathbf{g} denotes the first derivative of the objective function. The parameter β_k is important for nonlinear conjugate gradient methods, because it determines the type of method used. Since 1952, a series of representative conjugate gradient methods have been proposed, such as the Fletcher-Reeves (FR) [16], Polak-Ribère-Polyak (PRP) [17, 18], Hestens-Stiefel (HS) [15], Dai-Yuan (DY) [19], conjugate descent (CD) [20], and Liu-Storey (LS) [21] conjugate gradient methods. The common definitions of β_k are as follows.

$$\begin{aligned} \beta_k^{FR} &= \frac{\|\mathbf{g}_k\|^2}{\|\mathbf{g}_{k-1}\|^2}, \quad \beta_k^{PRP} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\|\mathbf{g}_{k-1}\|^2}, \\ \beta_k^{HS} &= \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}, \quad \beta_k^{DY} = \frac{\|\mathbf{g}_k\|^2}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}, \\ \beta_k^{CD} &= -\frac{\|\mathbf{g}_k\|^2}{\mathbf{d}_{k-1}^T \mathbf{g}_{k-1}}, \quad \beta_k^{LS} = -\frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{g}_{k-1}}, \end{aligned} \quad (7)$$

where $\|\cdot\|$ represents the Euclidean norm and $\mathbf{y}_{k-1} = \mathbf{g}_k - \mathbf{g}_{k-1}$.

In recent years, many scholars have improved the conjugate coefficient β for the conjugate gradient method [22] to achieve good convergence and numerical results [23, 24]. This study attempted the conjugate gradient modulation for gradient elements of the aperture gradient map. Only the basic conjugate gradient direction was used at this stage.

2. Search direction based on conjugate gradient modulation

The decrease in the objective function value of the column generation methods based on conjugate gradient directions with different coefficients in Eq. (7) was observed using the same objective function in a case of head and neck cancer (Fig. 1).

The objective function value of the column generation method based on the PRP conjugate gradient direction initially exhibited the fastest decline. During the second half of the iterative process, the objective function value of the column generation method based on the HS conjugate gradient direction exhibited the fastest convergence. Based on the above observations, an aperture shape optimization method, PRP-HS, based on modulation of the PRP and HS conjugate gradients, was proposed.

Two conjugate gradient descent directions were used to determine the search direction of the aperture shape. The

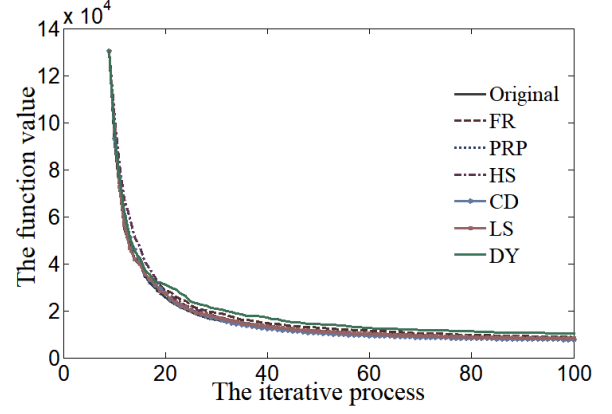


Fig. 1. Decrease in the objective function value of the above six column generation methods based on the conjugate gradient direction.

PRP conjugate gradient direction had priority for decision-making, which decreased as the optimization proceeded, and the HS conjugate gradient direction increased in decision-making priority. The expression for \mathbf{d}_k^{PRP-HS} is

$$\mathbf{d}_k^{PRP-HS} = \frac{1}{k} \mathbf{d}_k^{PRP} + \left(1 - \frac{1}{k}\right) \mathbf{d}_k^{HS}. \quad (8)$$

This formula can be rewritten as

$$\begin{aligned} \mathbf{d}_k^{PRP-HS} &= \frac{1}{k} (-\mathbf{g}_k + \beta_k^{PRP} \mathbf{d}_{k-1}^{PRP}) + \left(1 - \frac{1}{k}\right) (-\mathbf{g}_k + \beta_k^{HS} \mathbf{d}_{k-1}^{HS}) \\ &= -\mathbf{g}_k + \left(\frac{1}{k} \beta_k^{PRP} \mathbf{d}_{k-1}^{PRP} + \left(1 - \frac{1}{k}\right) \beta_k^{HS} \mathbf{d}_{k-1}^{HS}\right). \end{aligned} \quad (9)$$

The proposed method does not employ the hybrid conjugate gradient method [25] to generate the aperture shape. In contrast to the hybrid conjugate gradient method, the proposed method calculates and saves the corresponding \mathbf{d}_k^{PRP} and \mathbf{d}_k^{HS} . The direction \mathbf{d}_k^{PRP-HS} is obtained by combining two conjugate gradients through the number of iterations used to determine the aperture shape. The aperture search direction \mathbf{d}_k^{PRP-HS} is not used to calculate \mathbf{d}_{k+1}^{PRP} and \mathbf{d}_{k+1}^{HS} in the next iteration. In the proposed method, only basic conjugate gradient classes are used to modulate the gradient. According to the Eq. (8), to generate an aperture shape in each iteration, PRP conjugate gradient descent direction \mathbf{d}_k^{PRP} and HS conjugate gradient descent direction \mathbf{d}_k^{HS} should be calculated according to the aperture gradient information and previous direction information. With an increase in k , the search direction of the aperture shape \mathbf{d}_k^{PRP-HS} gradually moves from the PRP-based conjugate gradient descent direction to the HS-based conjugate gradient descent direction. When the optimization search is close to the optimal value $(\mathbf{D}_{js}^*, \mathbf{y}_{\kappa}^*)$, $\mathbf{g}_{k-1} \approx \mathbf{g}_k = 0$, according to Eq. (7), the values of β_k^{PRP} and β_k^{HS} are approximately zero. The condition that must be satisfied for the optimal solution is as follows:

$$\nabla_{D_{js}, y_{\kappa}} \mathcal{L}(D_{js}, y_{\kappa}, \pi_{js}, \rho_{\kappa}) = 0, \quad (10)$$

where $\mathcal{L}(x)$ is the Lagrange function and ρ_{κ} is another Lagrange multiplier under the KKT conditions [5]. At (D_{js}^*, y_{κ}^*) , subtract $\left(\frac{1}{k} \beta_{kjs}^{PRP} d_{(k-1)js}^{PRP} + (1 - \frac{1}{k}) \beta_{kjs}^{HS} d_{(k-1)js}^{HS}\right)$ from Eq. (10) without changing the optimal solution to the original optimization problem. The proposed method uses a conjugate gradient modulation to generate an aperture shape. Compared to generic column generation based on the negative gradient direction, the proposed method achieved faster convergence and shorter computation time for plan optimization.

III. EXPERIMENTAL SETUP AND EVALUATION CRITERIA

In this experiment, the classical pencil beam method [26] in an open-source computational environment for radiotherapy research (CERR) [27] was used to calculate the dose deposition matrix W . All methods involved in the experiment were implemented in Visual C++ (version Visual Studio 2012) on a computer with an Intel® Core™ i9-10900X central processing unit at 3.70 GHz, running on Windows 10 with 64 bits. All cancer cases involved in this study were obtained from Shanxi Provincial Cancer Hospital. The simulation experiments involving the relevant cancer cases in this study were conducted under a protocol approved by the Ethics Committee of the North University of China.

The physician defined the organs involved in the head and neck cancer cases, as illustrated in Fig. 2(a), based on image data. Three planning target volumes (PTVs), PTV 70 Gy, PTV 63 Gy, and PTV 56 Gy, were obtained by an outward expansion of 5 mm in each direction of the three clinical target volumes (CTVs). The ipsilateral parotid gland (IL-PG), contralateral parotid gland (CL-PG), spinal cord, and brain stem were considered as the organs at risk (OARs) in the optimization [28]. Those OARs were obtained by expanding 5 mm outwards from the outline of each organ. Nine equally spaced 6 MeV co-irradiated photon fields were set on the CERR to simulate the head and neck cancer case irradiation. In the optimization process, we constrained the dose distribution to the parotid glands using the dose-volume (DV) criterion [29] sub-objective function and to the spinal cord and brain stem using the maximum dose criterion to penalize doses beyond the upper limit, as listed in Table 1. The dose distributions to the PTVs were constrained using the mean and minimum dose sub-objective functions, and the dose distribution to the remaining normal tissues was constrained using the maximum dose sub-objective function. The optimization iterations were capped at 100 iterations.

For the prostate cancer cases illustrated in Fig. 2(b), one PTV was set, and the OARs were the bladder and rectum. The OARs and PTV were delineated by the physician using

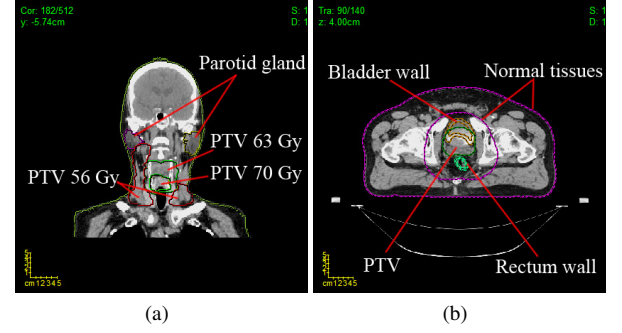


Fig. 2. Distribution of organs involved in the cancer cases: (a) for the head and neck cancer case and (b) for the prostate cancer case.

Table 1. DV constraint conditions for the head and neck cancer case.

Structure	DV parameter	
Parotid gland	$D_{mean} \leq 25 \text{ Gy}$	
Spinal cord	$D_{max} \leq 50 \text{ Gy}$	
Brain stem	$D_{max} \leq 54 \text{ Gy}$	
PTV 70 Gy	$V_{70 \text{ Gy}} > 95\%$	$V_{77 \text{ Gy}} < 5\%$
PTV 63 Gy	$V_{63 \text{ Gy}} > 95\%$	$V_{70 \text{ Gy}} < 5\%$
PTV 56 Gy	$V_{56 \text{ Gy}} > 95\%$	$V_{62 \text{ Gy}} < 5\%$

Table 2. DV constraint conditions for the prostate cancer case.

	Bladder	Rectum
DV parameter		$V_{50 \text{ Gy}} \leq 50\%$
		$V_{60 \text{ Gy}} \leq 35\%$
	$V_{65 \text{ Gy}} \leq 50\%$	$V_{65 \text{ Gy}} \leq 25\%$
	$V_{70 \text{ Gy}} \leq 35\%$	$V_{70 \text{ Gy}} \leq 20\%$
	$V_{75 \text{ Gy}} \leq 25\%$	$V_{75 \text{ Gy}} \leq 15\%$
	$V_{80 \text{ Gy}} \leq 15\%$	

the image data. Manually delineated bladder and rectal contours were expanded outwards by 5 mm to obtain the OARs. The CTV was expanded backward by 5 mm and outwards by 10 mm in the remaining directions to obtain the PTV. Five radioactive sources with gantry angles of 36°, 100°, 180°, 260°, and 324° were used for simulated irradiation. The dose distributions for the bladder and rectum were constrained using the DV criterion sub-objective function, and the DV constraint conditions are listed in Table 2. The mean and minimum dose criteria constrained the dose distribution to the PTV. The maximum dose sub-objective function restrained the remaining normal tissue dose distribution. The iterations of the optimization process were up to 60.

IV. RESULTS AND DISCUSSION

Comparative experiments were conducted to compare the performance of different methods for head and neck cancer cases (labeled as "H1," "H2," "H3," "H4," and "H5") and prostate cancer cases (labeled as "P1," "P2," "P3," "P4," and "P5"). The total objective function used in the experiments

was the sum of multiple sub-objective functions multiplied by the corresponding penalty factors [30]. In the same set of experiments, different contrast methods used the same objective function and penalty factors of the sub-objective functions. In Sec. IV A and Sec. IV B, the optimized results for cases H1, H2, P1, and P2 are presented. To evaluate the optimized results for cases H1, H2, P1, and P2, the DV histogram (DVH) was analyzed using the clinical guidance standard (Tables 1 and 2) developed by Marks et al. [31]. The generalized equivalent uniform dose (gEUD) and normal tissue complication probability (NTCP) of head and neck [32, 33] and prostate cancer cases [34, 35] were calculated to evaluate the protective effect of each method on the OARs. Relatively lower gEUD and NTCP values indicated better protection of the OARs. The conformity number (CN) [36] and homogeneity index (HI) [37] of the PTV were calculated. When CN and HI were close to 1, the dose distribution to the PTV was more conformal and uniform. The running time, number of apertures, and trend of the objective function during optimization were also used to investigate the performance of the experimental methods. The optimized results for the remaining six cancer cases are concisely presented.

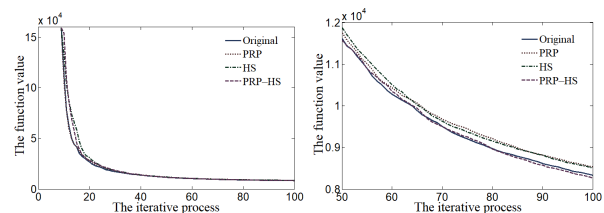
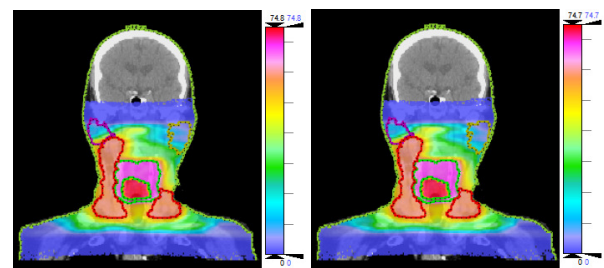
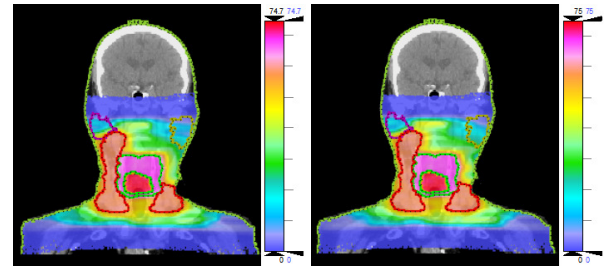
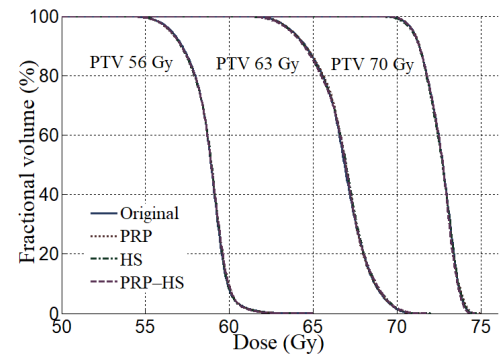
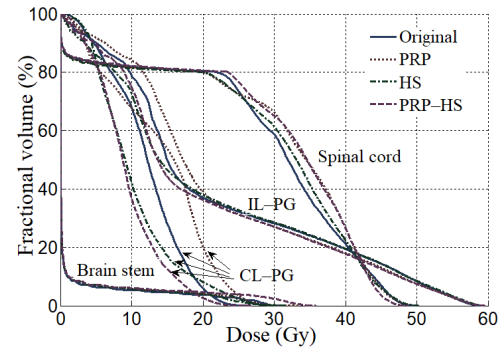


Fig. 3. Optimization results of case H1. (a) is the DVH of the OARs; (b) is the DVH of the PTVs after optimization; (c), (d), (e), and (f) are the dose distribution of case H1 on the same surfaces after optimization of the four methods; (g) is the decrease in the objective function value in the iterative process; and (h) is the enlarged figure of (g).

A. Results from cases of head and neck cancer

Four optimization methods—generic column generation (labeled as "Original"), PRP conjugate gradient direction-based column generation (labeled as "PRP"), HS conjugate gradient direction-based column generation (labeled as "HS"), and the proposed method integrating PRP and HS conjugate gradient directions (labeled as "PRP-HS")—were employed to optimize case H1. The optimization results are shown in Fig. 3 and Table 3.

The DV curves of the PTVs optimized by all the methods were mostly consistent (Fig. 3(b)). This conclusion can also be verified by the DV percentage, HI, and CN of the PTVs as shown in Table 3. It is apparent from the NTCP and gEUD of the OARs in Table 3 that PRP-HS can better protect the OARs while ensuring dose distribution to the PTVs. This conclusion is supported by the results shown in Figs. 3(c)–3(f). In particular, the parotid glands received the lowest dose after optimization of PRP-HS. Table 3, Fig. 3(g), and Fig. 3(h) also demonstrate that the optimization time of PRP-HS was the shortest, and the decrease in the objective function value was the largest in the optimization iteration process.

In this study, the generic aperture shape generation is regarded as a search for the negative-gradient descent direction. To address the shortcomings of the negative-gradient descent direction, the search direction was constructed using the conjugate gradient direction. However, when the classical conjugate gradient direction was used to construct the search direction, a decrease in the objective function value was not ideal for the entire iterative process. The proposed method based on conjugate gradient modulation was used to determine the aperture shape. The proposed method improves the search direction of generic column generation. Subsequent cases were optimized using only these two methods to verify the improved efficacy of PRP-HS over generic column generation. For H2, the DVHs, dose distributions, and decreases in the objective function values obtained using these two methods are shown in Fig. 4. The performance details of the methods for case H2 are presented in Table 4.

Similar to case H1, case H2 shows that the dose distributions of the PTVs were mostly consistent (Fig. 4(b) and Table 4). Table 4 shows that, compared with the Original, PRP-HS can reduce the NTCP and gEUD of the OARs. However, the optimized mean dose received by the parotid glands and the maximum dose received by the spinal cord and brain stem did not satisfy the evaluation criteria listed in Table 1. The dose slices shown in Figs. 4(c) and 4(d) show that the organ distribution in case H2 was more compact than that in case H1, which makes it more difficult for the dose on the OARs, after optimization, to meet the evaluation criteria in Table 1. Fig. 4(e) shows that, compared with the Original, PRP-HS can significantly reduce the objective function value.

B. Results from cases of prostate cancer

Two optimization methods, the Original and PRP-HS methods, were used to optimize cases P1 and P2. Fig. 5 depicts the optimization results, and Table 5 presents the evaluation index of the OARs and dose distribution to the PTV after optimization.

These two groups of comparative experiments verified the effectiveness of PRP-HS. In Figs. 5(a) and 5(b), PRP-HS reduced the DV curves of the OARs in the high-dose region, while the DV curves of the PTV were mostly consistent. The PTV indicators in Table 5 confirm a similar dose distribution to the PTV, reducing the NTCP and gEUD of the OARs. Although the improvement in the plan quality was not as obvious as the optimization results of cases H1 and H2, the objective function value decreased significantly during the iterative process (Figs. 5(c) and 5(d)).

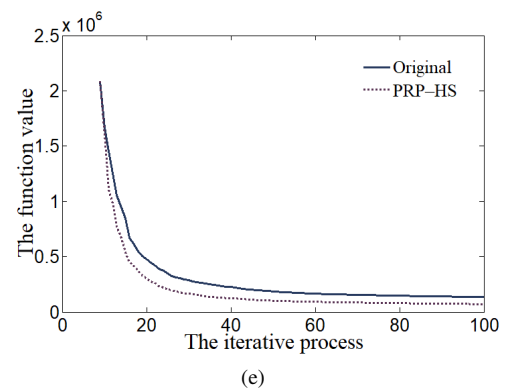
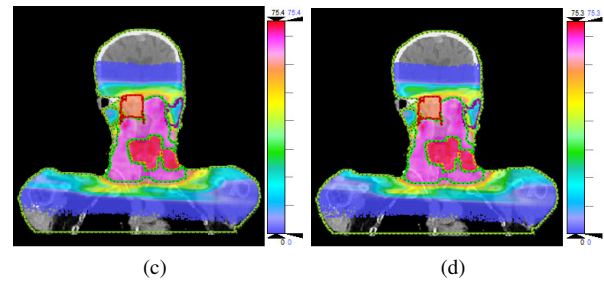
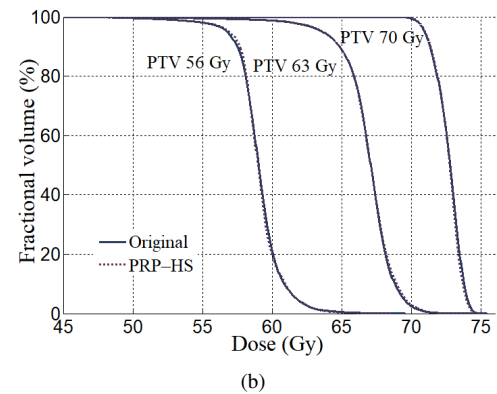
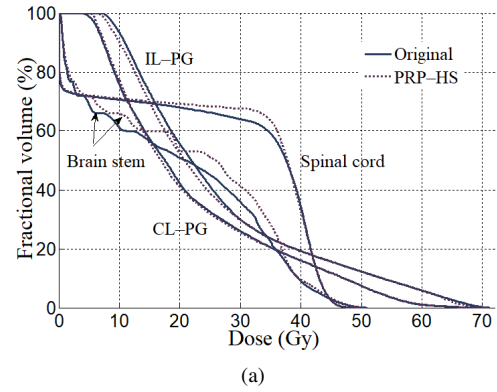


Fig. 4. Optimization results of case H2. (a) is the DVH of the OARs; (b) is the DVH of the PTVs; (c) and (d) are the dose distributions of case H2 on the same surfaces after optimization of the two methods; and (e) is the decrease in the objective function value in the iterative process.

Table 3. Results obtained by the four optimization methods for case H1.

		Original	PRP	HS	PRP-HS
PTV 70 Gy	V_{70} Gy (%)	99.3538	99.5830	99.1471	99.4928
	V_{77} Gy (%)	0	0	0	0
	HI	1.0409	1.0420	1.0423	1.0408
	CN	0.9474	0.9192	0.9318	0.9491
PTV 63 Gy	V_{63} Gy (%)	98.0154	97.9303	97.6194	97.6489
	V_{70} Gy (%)	1.5220	1.6016	1.9300	1.8231
	HI	1.0867	1.0871	1.0894	1.0897
	CN	0.2833	0.1876	0.1383	0.1585
PTV 56 Gy	V_{56} Gy (%)	97.6046	97.3125	97.3694	97.4009
	V_{62} Gy (%)	0.5494	0.5273	0.5122	0.4290
	HI	1.0658	1.0681	1.0668	1.0676
	CN	0.0092	0.0088	0.0073	0.0052
IL-PG	Mean dose (Gy)	21.8708	22.5257	21.1594	20.7717
	gEUD (Gy)	21.8708	22.5258	21.1593	20.7718
	NTCP (%)	10.08	15.53	7.83	6.78
CL-PG	Mean dose (Gy)	12.0590	14.0597	10.2722	9.3394
	gEUD (Gy)	12.0590	14.0599	10.2724	9.3394
	NTCP (%)	0.07	0.25	0.02	0.01
Spinal cord	Max dose (Gy)	49.7250	50.2250	50.2750	48.7750
	gEUD (Gy)	41.6845	41.7916	41.7300	40.9892
	NTCP (%)	1.65	1.69	1.66	1.42
Brain stem	Max dose (Gy)	30.9750	33.725	31.5750	35.8750
	gEUD (Gy)	15.9385	15.8683	15.1105	18.6819
	NTCP (%)	3.50E-06	3.35E-06	2.10E-06	1.79E-05
Aperture number		90	91	93	92
Time (s)		822.512	757.212	787.000	749.817

Table 4. Results obtained by the two optimization methods for case H2.

		Original	PRP-HS
PTV 70 Gy	V_{70} Gy (%)	99.4607	99.6216
	V_{77} Gy (%)	0	0
	HI	1.0433	1.0414
	CN	0.8442	0.8707
PTV 63 Gy	V_{63} Gy (%)	96.3363	96.4373
	V_{70} Gy (%)	2.5845	3.1131
	HI	1.0917	1.0931
	CN	0.0074	0.0115
PTV 56 Gy	V_{56} Gy (%)	97.0534	97.2243
	V_{62} Gy (%)	4.1580	4.1803
	HI	1.0880	1.0853
	CN	0.0006	0.0008
IL-PG	Mean dose (Gy)	26.6590	26.0172
	gEUD (Gy)	26.6585	26.0172
	NTCP (%)	36.67	32.06
CL-PG	Mean dose (Gy)	22.2407	22.1554
	gEUD (Gy)	22.2407	22.1552
	NTCP (%)	11.41	11.09
Spinal cord	Max dose (Gy)	50.8750	49.6750
	gEUD (Gy)	40.9685	40.9731
	NTCP (%)	1.41	1.41
Brain stem	Max dose (Gy)	50.9750	50.1250
	gEUD (Gy)	33.2127	33.9282
	NTCP (%)	0.02	0.03
Aperture number		85	85
Time (s)		865.506	861.881

C. Supplementary experiments

the proposed method.

The remaining cancer cases (labeled as "H3," "H4," "H5," "P3," "P4," and "P5") were used to verify the performance of

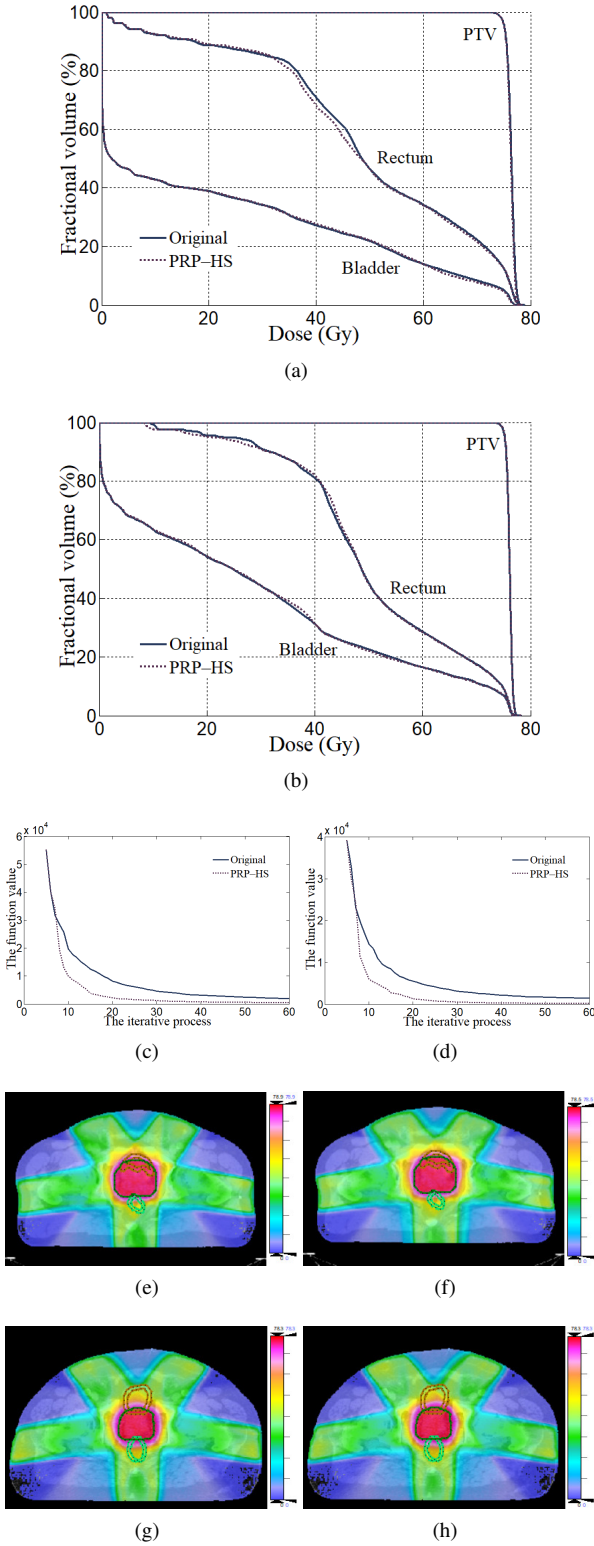


Fig. 5. Optimization results of cases P1 and P2. (a) is the DVH for case P1; (b) is the DVH for case P2; (c) and (d) are the decreases in the objective function in the iterative process for cases P1 and P2 respectively; (e) and (f) are the dose distributions of case P1 on the same surfaces after optimization of the two methods; and (g) and (h) are the dose distributions of case P2 on the same surfaces after optimization of the two methods.

On the premise that the dose distribution to the PTVs is generally consistent, Tables 6 and 7 present the optimization information for the OARs. For the three head and neck cancer cases in Table 6, under the premise of meeting the dose constraints in Table 1, the proposed method reduces the dose received by the parotid glands. As shown in Table 7, the proposed method reduced the NTCP and gEUD of the OARs in the three cases of prostate cancer.

D. Discussion

A suitable gradient descent direction was selected for optimization to solve the pricing problem in generic column generation. This is considered the direction of negative gradient descent. Although it has a low computational complexity, the steepest descent method is prone to the sawtooth effect when searching near the minimum value, resulting in reduced search efficiency. Compared with Newton's and quasi-Newton methods, the conjugate gradient method with its lower computational complexity is an ideal choice. Fig. 1 reveals that column generation methods based on conjugate gradient direction are superior to generic column generation in different degrees during the iteration. However, the convergence performance of column generation methods based on different conjugate gradient directions is not always suitable during the optimization process. A method based on conjugate gradient modulation was proposed to accelerate the descent speed of the objective function value without reducing the result quality. In Figs. 3(g) and 3(h), the proposed method combined the advantages of PRP and HS, and decreased faster than the Original method. The optimization times listed in Table 3 show that PRP-HS was the fastest. Compared to column generation based on the negative gradient direction, in this optimization process, the column generation method based on conjugate gradient modulation made the decrease in the objective function value more evident (Figs. 4(e), 5(c), and 5(d)).

When DVH is used for evaluation, the DV curves of the PTV optimized by all methods should be as similar as possible. Accordingly, the performance of the methods can be evaluated by observing the DVH of the OARs. For case H1, the DV curves of the PTVs in the DVH optimized using the four contrasting methods (Original, PRP, HS, and PRP-HS) were similar (Fig. 3(b)). The DVH of the OARs (Fig. 3(a)) showed that the DV curves obtained by PRP-HS were the lowest for the parotid glands. For the spinal cord and brain stem revealed that the DV curves optimized by PRP-HS were slightly worse than those obtained using the other three methods. However, the maximum doses received by the spinal cord and brain stem are greater concern (Table 1). The maximum dose received by the spinal cord optimized using PRP and HS exceeded 50 Gy, whereas when optimized using PRP-HS, the result was minimal (Table 3). The maximum dose received by the brain stem optimized using the four methods satisfied $D_{max} \leq 54$ Gy. For case H2, the gEUD of the spinal cord and brain stem optimized by PRP-HS were slightly worse than those obtained using the Original method

Table 5. Results obtained by the two optimization methods for cases P1 and P2.

		P1		P2	
		Original	PRP-HS	Original	PRP-HS
PTV	$V_{67.27 \text{ Gy}} (\%)$	100	100	100	100
	$V_{74 \text{ Gy}} (\%)$	99.2076	99.2488	99.6730	99.7988
	$V_{781.4 \text{ Gy}} (\%)$	0	0	0	0
	HI	1.0308	1.0289	1.0225	1.0215
	CN	0.9043	0.9222	0.9037	0.9044
Bladder	gEUD (Gy)	57.0297	56.6286	58.5413	58.4910
	NTCP (%)	5.09	4.79	6.38	6.34
Rectum	gEUD (Gy)	64.0253	63.8351	62.3576	62.3356
	NTCP (%)	10.03	9.73	7.60	7.57
Aperture number		57	58	58	58
Time (s)		360.064	345.325	377.920	361.984

Table 6. Results for cases H3, H4, and H5.

		H3		H4		H5	
		Original	PRP-HS	Original	PRP-HS	Original	PRP-HS
IL-PG	Mean dose (Gy)	32.8380	32.5670	40.3968	40.2397	35.4129	34.9677
CL-PG	Mean dose (Gy)	33.9796	31.6606	35.5173	34.0182	34.1495	33.2812
Spinal cord	Max dose (Gy)	49.1750	48.9250	49.7250	49.6250	49.2750	49.8250
Brain stem	Max dose (Gy)	51.9250	51.6250	50.6750	50.8750	50.1750	50.6250
Aperture number		90	89	83	84	92	94
Time (s)		1268.620	1232.551	802.318	760.827	1396.752	1265.277

Table 7. Results for cases P3, P4, and P5.

		P3		P4		P5	
		Original	PRP-HS	Original	PRP-HS	Original	PRP-HS
Bladder	gEUD (Gy)	66.4603	66.3782	61.8468	61.5174	49.7854	49.5949
	NTCP (%)	17.52	17.37	10.08	9.65	1.48	1.43
Rectum	gEUD (Gy)	62.916	62.6337	65.6279	65.5339	63.1659	62.8997
	NTCP (%)	8.36	7.97	12.86	12.68	8.72	8.34
Aperture number		59	59	57	57	58	60
Time (s)		262.975	237.437	252.116	246.576	309.672	292.155

Table 8. Statistical analysis of the optimization results.

			Original	PRP-HS	P-value
Head and neck cancer	IL-PG	Mean dose (Gy)	31.44±7.29	30.91±7.63	0.034
	CL-PG	Mean dose (Gy)	27.59±10.20	26.09±10.51	0.035
	Spinal cord	Max dose (Gy)	49.76±0.68	49.37±0.48	0.028
	Brain stem	Max dose (Gy)	46.95±8.95	47.83±6.70	0.441
Prostate cancer	Bladder	gEUD (Gy)	58.73±6.17	58.52±6.20	0.037
		NTCP (%)	8.11±6.09	7.92±6.06	0.062
	Rectum	gEUD (Gy)	63.62±1.27	63.45±1.29	0.027
		NTCP (%)	9.51±2.07	9.26±2.08	0.020

(Table 4). By definition, the gEUD examines the entire dose distribution in the whole organ. As shown in Fig. 4(a), the overall DV curves trend of the spinal cord and brain stem obtained with PRP-HS was worse than that of the Original method. According to Table 4, the maximum dose received by the brain stem optimized using both optimization methods satisfies $D_{max} \leq 54$ Gy. The maximum dose received by the spinal cord, optimized by the Original method, exceeded 50 Gy. Compared with the contrast method, PRP-HS reduced the NTCP and gEUD of the parotid glands to varying degrees and had a better ability to protect the OARs. These results illustrate that the proposed optimization method can better protect the OARs while ensuring dose distribution to the PTVs than the generic method. The optimization results for P1 and P2 (Fig. 5 and Table 5) also prove the proposed method performance.

For cases of head and neck cancer, only the evaluation index in Table 1 is presented in Table 6, whereas for cases of prostate cancer, only the NTCP and gEUD of the OARs are presented in Table 7. These results are sufficient to illustrate

the improvement in the optimization for the proposed method compared with the generic method.

Table 8 presents the optimization results for the five cases of head and neck cancer, and the five cases of prostate cancer involved in this study were statistically analyzed. The proposed method significantly improved the dose distribution to the OARs ($P < 0.05$) other than the brain stem ($P = 0.441$), and the stability of the proposed method was proven.

Compared to generic column generation based on negative gradient direction, PRP-HS required less time to optimize the treatment plan. According to the analysis of the optimization results, compared with generic column generation, the proposed optimization method did not reduce and, in some cases, improved the plan quality. The proposed method did not reduce the number of generated apertures, which is a direction for future research and improvement. In the DAO, the aperture shape can be optimized by selecting the descending direction that maximizes improving the objective function. To generate a deliverable aperture shape, the descent direction was the direction after the conjugate gradient modulation in the pricing problem of this study. The proposed aperture shape optimization method does not involve line search or step length selection. In the proposed method, only basic conjugate gradient classes were used to modulate the gradient. In the future, some newly proposed conjugate gradient methods

and radiation therapy techniques [38, 39] will be introduced into the DAO.

V. CONCLUSION

This study proposed a method based on conjugate gradient modulation for aperture shape optimization by modulating the gradients, because the negative gradient direction search exhibited a sawtooth effect when approaching the local optimal point. The conjugate gradient method with low computational cost and fast convergence was introduced into the aperture generation. The performance of PRP-HS was verified in head and neck cancer cases, and prostate cancer cases. Based on comparative experimental results, the proposed optimization method can accelerate the solution process and improve the quality of the treatment plan. The optimization time of the proposed method decreased by up to 9.41% for head and neck cancer cases, and 9.71% for prostate cancer cases. This improvement does not come at the expense of the quality of the results. While ensuring dose distribution to the PTV, PRP-HS reduced NTCP by up to 4.61% compared with generic column generation. According to the experimental results, the proposed aperture shape optimization method can be applied to radiotherapy plan optimization for different cancer cases and can be efficiently used in clinical settings.

- [1] D. M. Shepard, M. A. Earl, X. A. Li, *et al.* Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. *Med. Phys.* **29**, 1007-1018 (2002). DOI: [10.1118/1.1477415](https://doi.org/10.1118/1.1477415)
- [2] Y. J. Li, J. Yao, D. Z. Yao, Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy. *Phys. Med. Biol.* **48**, 3353-3374 (2003). DOI: [10.1088/0031-9155/48/20/007](https://doi.org/10.1088/0031-9155/48/20/007)
- [3] M. A. Earl, M. K. N. Afghan, C. X. Yu, *et al.* Jaws-only IMRT using direct aperture optimization. *Med. Phys.* **34**, 307-314 (2007). DOI: [10.1118/1.2403966](https://doi.org/10.1118/1.2403966)
- [4] B. Hårdemark, A. Liander, H. Reh binder, *et al.* Direct machine parameter optimization with ray machine in pinnacle. Raysearch Lab, Sweden, White Pape (2003).
- [5] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, *et al.* A column generation approach to radiation therapy treatment planning using aperture modulation. *SIAM J. Optimiz.* **15**, 838-862 (2005). DOI: [10.1137/040606612](https://doi.org/10.1137/040606612)
- [6] F. Preciado-Walters, M. P. Langer, R. L. Rardin, *et al.* Column generation for IMRT cancer therapy optimization with implementable segments. *Ann. Oper. Res.* **148**, 65-79 (2006). DOI: [10.1007/s10479-006-0080-1](https://doi.org/10.1007/s10479-006-0080-1)
- [7] Z. F. Dai, T. Y. Li, M. Yang, Forecasting stock return volatility: The role of shrinkage approaches in a data-rich environment. *Journal of Forecasting* **41**, 980-996 (2022). DOI: [10.1002/for.2841](https://doi.org/10.1002/for.2841)
- [8] Z. F. Dai, H. Y. Zhu, Dynamic risk spillover among crude oil, economic policy uncertainty and Chinese financial sectors. *Int. Rev. Econ. Financ.* **83**, 421-450 (2023). DOI: [10.1016/j.iref.2022.09.005](https://doi.org/10.1016/j.iref.2022.09.005)
- [9] Z. F. Dai, X. T. Zhang, Climate policy uncertainty and risks taken by the bank: Evidence from China. *Int. Rev. Financ. Anal.* **87**, 102579 (2023). DOI: [10.1016/j.irfa.2023.102579](https://doi.org/10.1016/j.irfa.2023.102579)
- [10] R. Cao, X. Pei, H. Zheng, *et al.* Direct aperture optimization based on genetic algorithm and conjugate gradient in intensity modulated radiation therapy. *Chin. Med. J.* **127**, 4152-4153 (2014). DOI: [10.3760/cma.j.issn.0366-6999.20130644](https://doi.org/10.3760/cma.j.issn.0366-6999.20130644)
- [11] F. Carlsson, A. Forsgren, On column generation approaches for approximate solutions of quadratic programs in intensity-modulated radiation therapy. *Ann. Oper. Res.* **223**, 471-481 (2014). DOI: [10.1007/s10479-013-1360-1](https://doi.org/10.1007/s10479-013-1360-1)
- [12] R. H. Byrd, P. Lu, J. Nocedal, *et al.* A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16**, 1190-1208 (1995). DOI: [10.1137/0916069](https://doi.org/10.1137/0916069)
- [13] C. Zhu, R. H. Byrd, P. Lu, *et al.* Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM T. Math. Software* **23**, 550-560 (1997). DOI: [10.1145/279232.279236](https://doi.org/10.1145/279232.279236)
- [14] J. L. Morales, J. Nocedal, Remark on "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization". *ACM Transactions on Mathematical Software* **38**, 1-4 (2011). DOI: [10.1145/2049662.2049669](https://doi.org/10.1145/2049662.2049669)
- [15] M. R. Hestenes, E. L. Stiefel, Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards* **49**, 409-436 (1952). DOI: [10.6028/jres.049.044](https://doi.org/10.6028/jres.049.044)
- [16] R. Fletcher, C. M. Reeves, Function minimization by conjugate gradients. *The Computer Journal* **7**, 149-154 (1964). DOI: [10.1093/comjnl/7.2.149](https://doi.org/10.1093/comjnl/7.2.149)
- [17] E. Polak, G. Ribiere, Note sur la convergence des methodes de directions conjuguees. *Rev Francaise Informat Recherche Operationelle* **3**, 35-43 (1969). DOI: [10.1051/m2an/196903r100351](https://doi.org/10.1051/m2an/196903r100351)

- [18] B. T. Polyak, The conjugate gradient method in extreme problems. *USSR Computational Mathematics and Mathematical Physics* **9**, 94-112 (1969). DOI: [10.1016/0041-5553\(69\)90035-4](https://doi.org/10.1016/0041-5553(69)90035-4)
- [19] Y. H. Dai, Y. Yuan, A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property. *SIAM J. Optimiz.* **10**, 177-182 (1999). DOI: [10.1137/S1052623497318992](https://doi.org/10.1137/S1052623497318992)
- [20] R. Fletcher, in *Practical Methods of Optimization*. Unconstrained Optimization, Vol. 1 (John Wiley and Sons Press, New York, 1987).
- [21] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms, Part 1: Theory. *J. Optimiz. Theory App.* **69**, 129-137 (1991). DOI: doi.org/10.1007/bf00940464
- [22] W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods. *Pac. J. Optim.* **2**, 35-58 (2006).
- [23] L. Zhang, W. Zhou, D. Li, A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence. *IMA J. Numer. Anal.* **26**, 629-640 (2006). DOI: [10.1093/imanum/drl016](https://doi.org/10.1093/imanum/drl016)
- [24] M. Li, H. Liu, Z. Liu, A new family of conjugate gradient methods for unconstrained optimization. *J. Appl. Math. Comput.* **58**, 219-234 (2018). DOI: [10.1007/s12190-017-1141-0](https://doi.org/10.1007/s12190-017-1141-0)
- [25] P. Mtgulwa, P. Kaelo, An efficient modified PRP-FR hybrid conjugate gradient method for solving unconstrained optimization problems. *Appl. Numer. Math.* **145**, 111-120 (2019). DOI: [10.1016/j.apnum.2019.06.003](https://doi.org/10.1016/j.apnum.2019.06.003)
- [26] A. Ahnesjö, M. Saxner, A. Trepp, A pencil beam model for photon dose calculation. *Med. Phys.* **19**, 263-273 (1992). DOI: [10.1118/1.596856](https://doi.org/10.1118/1.596856)
- [27] J. O. Deasy, A. I. Blanco, V. H. Clark, CERR: a computational environment for radiotherapy research. *Med. Phys.* **30**, 979-985 (2003). DOI: [10.1118/1.1568978](https://doi.org/10.1118/1.1568978)
- [28] G. Mu, E. Ludlum, P. Xia, Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer. *Phys. Med. Biol.* **53**, 77-88 (2008). DOI: [10.1088/0031-9155/53/1/005](https://doi.org/10.1088/0031-9155/53/1/005)
- [29] Q. Wu, R. Mohan, Algorithms and functionality of an intensity modulated radiotherapy optimization system. *Med. Phys.* **27**, 701-711 (2000). DOI: [10.1118/1.598932](https://doi.org/10.1118/1.598932)
- [30] M. L. Kessler, D. L. Mcshan, M. A. Epelman, *et al.* Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans. *Optim. Eng.* **6**, 421-448 (2005). DOI: [10.1007/s11081-005-2066-2](https://doi.org/10.1007/s11081-005-2066-2)
- [31] L. B. Marks, E. D. Yorke, A. Jackson, *et al.* Use of normal tissue complication probability models in the clinic. *Int. J. Radiat. Oncol.* **76**, S10-S19 (2010). DOI: [10.1016/j.ijrobp.2009.07.1754](https://doi.org/10.1016/j.ijrobp.2009.07.1754)
- [32] A. Eisbruch, R. K. Ten Haken, H. M. Kim, *et al.* Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. *Int. J. Radiat. Oncol.* **45**, 577-587 (1999). DOI: [10.1016/S0360-3016\(99\)00247-3](https://doi.org/10.1016/S0360-3016(99)00247-3)
- [33] C. Burman, G. J. Kutcher, B. Emami, *et al.* Fitting of normal tissue tolerance data to an analytic function. *Int. J. Radiat. Oncol.* **21**, 123-135 (1991). DOI: [10.1016/0360-3016\(91\)90172-Z](https://doi.org/10.1016/0360-3016(91)90172-Z)
- [34] E. Dale, T. P. Hellebust, A. Skjønberg, *et al.* Modeling normal tissue complication probability from repetitive computed tomography scans during fractionated high-dose-rate brachytherapy and external beam radiotherapy of the uterine cervix. *Int. J. Radiat. Oncol.* **47**, 963-971 (2000). DOI: [10.1016/S0360-3016\(00\)00510-1](https://doi.org/10.1016/S0360-3016(00)00510-1)
- [35] S. T. H. Peeters, M. S. Hoogeman, W. D. Heemsbergen, *et al.* Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: Normal tissue complication probability modeling. *Int. J. Radiat. Oncol.* **66**, 11-19 (2006). DOI: [10.1016/j.ijrobp.2006.03.034](https://doi.org/10.1016/j.ijrobp.2006.03.034)
- [36] A. van't Riet, A. C. Mak, M. A. Moerland, *et al.* A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: Application to the prostate. *Int. J. Radiat. Oncol.* **37**, 731-736 (1997). DOI: [10.1016/S0360-3016\(96\)00601-3](https://doi.org/10.1016/S0360-3016(96)00601-3)
- [37] N. Hodapp, The ICRU report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). *Strahlenther Onkol.* **188**, 97-99 (2012). DOI: [10.1007/s00066-011-0015-x](https://doi.org/10.1007/s00066-011-0015-x)
- [38] Y. Luo, S. C. Huang, H. Zhang, *et al.* Assessment of the induced radioactivity in the treatment room of the heavy-ion medical machine in Wuwei using PHITS. *Nucl. Sci. Tech.* **34**, 29 (2023). DOI: [10.1007/s41365-023-01181-8](https://doi.org/10.1007/s41365-023-01181-8)
- [39] Y. Q. Yang, W. C. Fang, X. X. Huang, *et al.* Static superconducting gantry-based proton CT combined with X-ray CT as prior image for FLASH proton therapy. *Nucl. Sci. Tech.* **34**, 11 (2023). DOI: [10.1007/s41365-022-01163-2](https://doi.org/10.1007/s41365-022-01163-2)